

Model Formulation ■

HL7 Clinical Document Architecture, Release 2

ROBERT H. DOLIN, MD, LIORA ALSCHULER, SANDY BOYER, BSP, CALVIN BEEBE,
FRED M. BEHLEN, PhD, PAUL V. BIRON, AMNON SHABO (SHVO), PhD

Abstract Clinical Document Architecture, Release One (CDA R1), became an American National Standards Institute (ANSI)-approved HL7 Standard in November 2000, representing the first specification derived from the Health Level 7 (HL7) Reference Information Model (RIM). CDA, Release Two (CDA R2), became an ANSI-approved HL7 Standard in May 2005 and is the subject of this article, where the focus is primarily on how the standard has evolved since CDA R1, particularly in the area of semantic representation of clinical events. CDA is a document markup standard that specifies the structure and semantics of a clinical document (such as a discharge summary or progress note) for the purpose of exchange. A CDA document is a defined and complete information object that can include text, images, sounds, and other multimedia content. It can be transferred within a message and can exist independently, outside the transferring message. CDA documents are encoded in Extensible Markup Language (XML), and they derive their machine processable meaning from the RIM, coupled with terminology. The CDA R2 model is richly expressive, enabling the formal representation of clinical statements (such as observations, medication administrations, and adverse events) such that they can be interpreted and acted upon by a computer. On the other hand, CDA R2 offers a low bar for adoption, providing a mechanism for simply wrapping a non-XML document with the CDA header or for creating a document with a structured header and sections containing only narrative content. The intent is to facilitate widespread adoption, while providing a mechanism for incremental semantic interoperability.

■ J Am Med Inform Assoc. 2006;13:30–39. DOI 10.1197/jamia.M1888.

Clinical Document Architecture, Release One (CDA R1), became an American National Standards Institute (ANSI)-approved Health Level 7 (HL7) Standard in November 2000, representing the first specification derived from the HL7 Reference Information Model (RIM).^{1,2} CDA, Release Two (CDA R2), became an ANSI-approved HL7 standard in May 2005³ and is the subject of this article, where the focus

is primarily on how the standard has evolved since CDA R1, particularly in the area of semantic representation of clinical events.

The basic model of CDA R2 is essentially unchanged from CDA R1. A CDA document has a header and a body. The header identifies and classifies the document and provides information on authentication, the encounter, the patient, and the involved providers. The body contains the clinical report, organized into sections whose narrative content can be encoded using standard vocabularies. The main evolutionary step in CDA R2 is that, whereas CDA R1 only used the RIM to derive the header, in CDA R2 both header and body are fully RIM derived. CDA R2 enables clinical content in the document body to be formally expressed to the extent that it is modeled in the RIM, coupled with terminology.

CDA R1 is in use worldwide, and many of these sites are already making the transition to CDA R2.^{4–7} Both Finland and Greece expect to have the majority of their populations' health records accessible in national information infrastructures within the next two or three years where they have been using CDA documents since the release of CDA R1. The cost-effective and rapid proliferation of accessibility to clinical information in these countries relies heavily on CDA documents created from Web-based, small office electronic health records and from legacy hospital information systems. Several sites have developed decision support applications using CDA R1 with local extensions or prenormative drafts

Affiliations of the authors: Kaiser Permanente, Pasadena, CA (RHD, PVB); Alschuler Associates, LLC, East Thetford, VT (LA); Consultant, Laguna Beach, CA (SB); Mayo Clinic, Rochester, MN (CB); LAI Technology, Homewood, IL (FMB); Amnon Shabo (Shvo), IBM Haifa Research Lab, Haifa, Israel (AS).

It is impossible to acknowledge all those whose foundational work has led to the development of the HL7 Development Framework and the HL7 Reference Information Model, from which CDA R2 derives. That CDA R2 is even possible is thanks to the years of work that have gone into the creation and consensus around these foundational components.

The authors acknowledge and thank the HL7 organization and membership for providing challenging use cases, for striving toward greater and greater semantic interoperability, and for creating an atmosphere of camaraderie where people work together to solve real health care challenges that benefit from interoperability solutions.

Correspondence and reprints: Robert H. Dolin, MD, 411 N. Lakeview Avenue, Anaheim, CA 92807; e-mail: <robert.h.dolin@kp.org>.

Received for review: 06/12/05; accepted for publication: 09/20/05.

of CDA R2, and CDA has already outlived changes in communication and access methodology. Within the United States, CDA is cited in the plans for the emerging information exchange networks and is the basis for the planned Health Insurance Portability and Accountability Act (HIPAA)-compliant claims, referrals, and authorization processes.^{8,9} The flagship implementation for CDA in the United States remains the Mayo Clinic where the Notes project will ultimately scale to 50,000 notes per week (Calvin Beebe, personal communication). Several academic medical centers are working with CDA.¹⁰

The overriding driving force behind the development of CDA R2 has been the desire to further encode the narrative clinical statements found in clinical reports, and to do so in such a way as to enable comparison of content from documents created on information systems of widely varying characteristics. Requirements used to guide the process have come from the implementations noted above, comparison with other models (such as CEN ENV 13606, *openEHR*, and DICOM¹¹⁻¹³), national initiatives for exchange of referral and medical summary documents, trade show demonstrations, medical natural language processing models,¹⁴ comparison with other HL7 V3 specifications, and more. While CDA R2 does not fully enable plug and play semantic interoperability, it takes us yet another step closer.

This article is an introduction and overview to CDA R2. It is geared toward medical informaticians who do not have significant familiarity with HL7 Version 3, and is intended to introduce the approach and objectives used in the creation of the standard and present an overview of the standard, not sufficient for implementation but sufficiently detailed to enable the reader to understand the scope and contents of the standard. Interested readers looking for detailed descriptions are encouraged to contact HL7 (www.hl7.org) for a copy of the normative specification.

CDA R2 Overview

What Is the CDA?

The HL7 CDA is a document markup standard that specifies the structure and semantics of a clinical document (such as a discharge summary, progress note, procedure report) for the purpose of exchange. A CDA document is a defined and complete information object that can include text, images, sounds, and other multimedia content. It can be transferred within a message, and can exist independently, outside the transferring message.

CDA documents are encoded in Extensible Markup Language (XML).¹⁵ They derive their machine processable meaning from the HL7 RIM¹⁶ and use the HL7 Version 3 data types. The RIM and the V3 data types provide a powerful mechanism for enabling CDA's incorporation of concepts from standard coding systems such as Systemized Nomenclature of Medicine Clinical Terms (SNOMED CT) and Logical Observation Identifiers Names and Codes (LOINC).¹⁷⁻¹⁹

The CDA specification is richly expressive and flexible and is designed to be broad enough to cover the domain of clinical documents. Templates and/or implementation guides can be used to constrain the CDA specification within a particular implementation and to provide validating rule sets that check conformance to these constraints.

Major Components of a CDA Document

Major components of a prototypic CDA document are shown in Figure 1. Note that many required components are left out to simplify the example.

A CDA document is wrapped by the <ClinicalDocument> element and contains a header and a body. The header lies between the <ClinicalDocument> and the <structuredBody> elements and identifies and classifies the document and provides information on authentication, the encounter, the patient, and the involved providers.

The body contains the clinical report and can be either an unstructured blob or can be comprised of structured markup. Figure 1 shows a structured body, which is wrapped by the <structuredBody> element and is divided up into recursively nestable document sections.

A CDA document section is wrapped by the <section> element. Each section can contain a single "narrative block" and any number of CDA entries and external references. The narrative block is a critical component of CDA and must contain the human readable content to be rendered. It is wrapped by the <text> element within the <section> element and contains XML markup that is similar to XHTML.²⁰ The "originator" (defined as the application role responsible for creation of a conformant CDA document) must ensure that the attested portion of the document body is conveyed in narrative blocks such that a recipient, adhering to recipient rendering rules, will correctly render the document. This process ensures human readability and enables a recipient to receive a CDA document from anyone and faithfully render the attested content using a single style sheet.

Within a document section, the narrative block represents content to be rendered, whereas CDA entries represent structured content provided for further computer processing (e.g., decision-support applications). CDA entries typically encode content present in the narrative block of the same section. Figure 1 shows two <observation> CDA entries and a <substanceAdministration> entry containing a nested <supply> entry, although several other CDA entries are defined. These entries are derived from classes in the RIM and enable formal representation of clinical statements in the narrative.

```
<ClinicalDocument>
... CDA Header ...
<structuredBody>
  <section>
    <text>(a.k.a. "narrative block")</text>
    <observation>...</observation>
    <substanceAdministration>
      <supply>...</supply>
    </substanceAdministration>
    <observation>
      <externalObservation>...
    </externalObservation>
    </observation>
  </section>
  <section>
    <section>...</section>
  </section>
</structuredBody>
</ClinicalDocument>
```

Figure 1. Major components of a Clinical Document Architecture (CDA) document.

While the narrative blocks must always be present, the CDA entries are optional. An originator of a CDA document is not required to fully encode all narrative into CDA entries within the CDA body, nor is a recipient required to parse and interpret the complete set of CDA entries contained within the CDA body. Within an implementation, trading partners may ascribe additional originator and recipient responsibilities to create various entries and may create various templates and/or implementation guides that require the use of various entries. As a result, CDA R2 can be relatively simple to implement (i.e., just narrative blocks) or can be relatively detailed to implement (i.e., with the inclusion of many rich and expressive entries) and provides a migration pathway toward progressively richer computer-processable content.

CDA entries can nest and can reference external objects. CDA external references always occur within the context of a CDA entry. External references refer to content that exists outside the CDA document, such as some other image, some other procedure, or some other observation (which is wrapped by the `<externalObservation>` element in Figure 1).

CDA Documents and HL7 Messages

CDA specifies the structure and semantics of clinical documents (such as discharge summaries, progress notes) for the purpose of exchange. The scope has expanded somewhat through usage (for instance, some implementations are using CDA to exchange laboratory reports or prescriptions), and a common question relates to the distinction between an HL7 document and an HL7 message, and knowing which to use when. While there are gray zones, messages tend to be transient, trigger based, and nonpersistent, whereas clinical documents have persistence, wholeness, and clinician authentication and are human readable.

CDA documents can be exchanged in HL7 messages or exchanged using other transport solutions. The exact method used is outside the scope of the standard, but must take a number of requirements into account: All components of a CDA document that are integral to its state of wholeness (such as attested multimedia) can be exchanged as a unit; content needing to be rendered or additional files associated with a CDA document (such as a style sheet) can be included in the exchange package; there is no need to change any of the references (e.g., a reference to attested multimedia in a separate file) within the base CDA document when creating or extracting the exchange package (indeed, they cannot be changed); there are no restrictions on the directory structure used by receivers—receivers can place the components of the CDA document into directories of their choosing; critical metadata about the CDA instance needed for document management (e.g., document state, document archival status) must be included in the exchange package.

CDA R2 recommends the use of RFC 2557 “MIME Encapsulation of Aggregate Documents, such as HTML (MHTML),”²¹ which provides a standards-based solution to packaging the CDA document and associated files, inside of a message designed to carry documents, such as the HL7 V2 or V3 Medical Records message.

CDA Extensibility

Locally defined markup can be used to extend CDA when local semantics have no corresponding representation in the CDA specification. CDA seeks to standardize the highest

level of shared meaning while providing a clean and standard mechanism for tagging meaning that is not shared. To support local extensibility requirements, it is permitted to include additional XML elements and attributes that are not included in the CDA schema. These extensions should not change the meaning of any of the standard data items, and receivers must be able to safely ignore these elements. Document recipients must be able to faithfully render the CDA document while ignoring extensions. When these extension mechanisms mark up content of general relevance, HL7 encourages users to get their requirements formalized in a subsequent version of the standard so as to maximize the use of shared semantics.

CDA R2 Technical Artifacts

Several articles have described the HL7 Version 3 development process, whereby the central HL7 RIM, which is the definitive source for definitions of components used in all HL7 V3 specifications, is used to derive a formal specification.^{22–24}

The definitive reference describing the methodology is the HL7 Development Framework (HDF),²⁵ which governs the process used to derive the CDA R2 object model from the RIM.

The CDA schema is derived from the CDA R2 object model per the HL7 XML Implementation Technology Specification, which defines the XML conventions used by HL7 V3 specifications and algorithmically converts the object model into an XML schema.^{26,27} The CDA narrative block markup, which is the XML content model of Section.text, is manually crafted (i.e., not RIM derived) and represents an extended subset of XHTML²⁰ that is backwardly compatible with CDA R1.

The following sections describe the technical artifacts that result from application of the HDF. While many implementers are primarily interested in the CDA R2 XML schema, they will often refer to the object model to gain a deeper understanding of various aspects of the standard.

HL7 V3 Data Types and Vocabulary

Data types define the structural format of the data carried within an RIM class’s attribute and influence the set of allowable values an attribute may assume. Some data types have very little intrinsic semantic content (e.g., INTEGER, TIME STAMP). However, HL7 also defines more extensive data types such as GENERAL TIMING SPECIFICATION (GTS), which supports complex temporal expressions, and CONCEPT DESCRIPTOR (CD), which supports the postcoordination of codes (or, stated in another way, the combining of codes from a terminology to create a more specific concept).

Several CDA components (e.g., ClinicalDocument.code and SubstanceAdministration.routeCode, further described below) are designed to carry concepts drawn from HL7-defined or HL7-recognized coding systems such as LOINC or SNOMED CT. Value sets for these components specify allowable codes, and each value set is assigned a coding strength of “Coded, No Extensions” (CNE), in which case, the only allowable values for the CDA component are those in the stated value set; or “Coded, With Extensions” (CWE), in which case values outside the stated value set can be used if necessary.

Post-coordination is allowed in CDA components (such as Observation.code, further described below) that use the CD data type. For example, SNOMED CT defines a concept “cellulitis,” an attribute “finding site,” and a concept “foot structure,” which can be combined in Observation.code to create a post-coordinated expression. Alternatively, Observation.code

can be valued with the single SNOMED CT concept “cellulitis of foot.” While the CD data type provides syntax for combining multiple codes, it is the source terminology that specifies the rules for what codes can be combined, and the semantics for aggregating pre- and post-coordinated expressions.^{28,29}

CDA R2 Object Model

The CDA R2 object model is a technical diagram of the CDA specification. It is presented using conventions and notations that were developed by HL7 to represent the specific semantic constructs of the RIM. Although it could be represented in Unified Modeling Language (UML)³⁰ notation, as the RIM is, the HL7 notation provides more details about the specific constraints and class refinements being represented. The HL7 diagramming convention abbreviates some relationships, enabling diagrams to be smaller and more concise and to convey more information visually.

Portions of the CDA R2 object model are shown in Figures 2 and 3. Figure 2 shows a portion of the header and its connection to the document body and document sections. Figure 3 shows the connection from a document section to nested CDA entries. These components are further described in the sections that follow.

Colors in the figures correspond to the RIM area from which the class is derived. The RIM is a consensus-based model of the health care domain, composed of six “back-bone” classes (Act, Participation, Entity, Role, ActRelationship, RoleLink) and their specializations. “Act” represents the actions that are executed and must be documented as health care is managed and provided (e.g., appendectomy; serum sodium measurement); “Participation” expresses the context for an act such as who performed it, for whom it was done, where it

was done (e.g., author, attending physician); “Entity” represents the physical things and beings that are of interest to and take part in health care (e.g., person, organization); “Role” establishes the roles that entities play as they participate in health care acts (e.g., patient, family member); “ActRelationship” represents a relationship between two Acts (e.g., causality, indication); and “RoleLink” represents a dependency between two Roles (e.g., one role has authority over another role). Act specializations are colored red, Participations are colored blue, Entities are green, Roles are yellow, and Act Relationships are pink.

Every Act has a mandatory moodCode attribute, which distinguishes the Act as a factual statement or in some other manner such as a command, possibility, and goal. The (CNE) HL7-defined value set for moodCode includes “EVN” (event), meaning that the Act is describing something that occurred; “DEF” (definition), meaning that the Act is providing a master file type of description; “INT” (intent), where the Act is describing an action plan or order; “GOL” (goal), for describing a desired outcome; “EVN.CRT” (event criterion), which represents an event that must apply for an associated service to be considered.

An Act can have zero to many ActRelationships to other Acts and can have zero to many Participations, each played by an Entity in some Role. A Role is a relationship between two Entities, the Entity playing the Role (represented as a solid line between Role and Entity), and the Entity who recognizes or assigned the role (represented as a dashed line between Role and Entity). Thus, in Figure 2, a “legalAuthenticator” is a Participant of a “ClinicalDocument” Act and is played by a “Person” Entity in an “AssignedEntity” Role that is recognized by an “Organization” Entity.

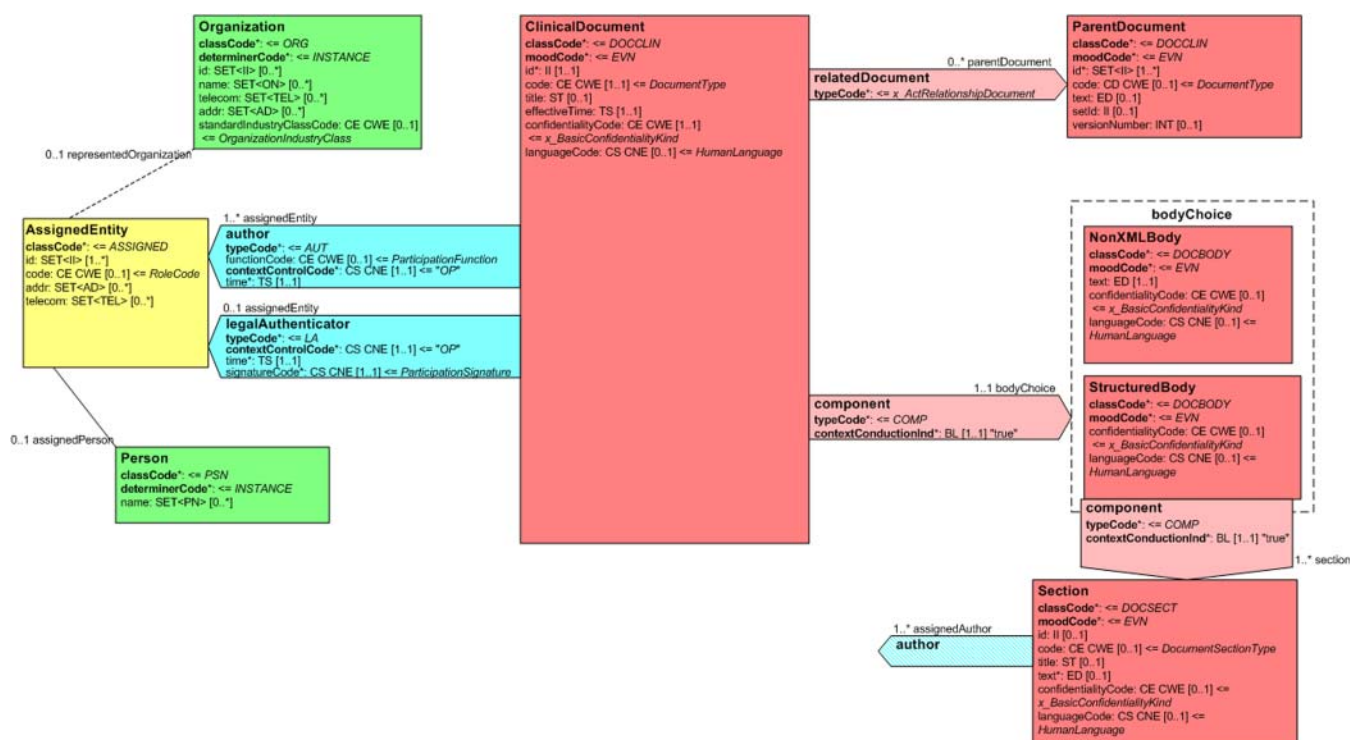


Figure 2. Clinical Document Architecture (CDA) Release 2 object model showing a portion of the header and its connection to the document body.

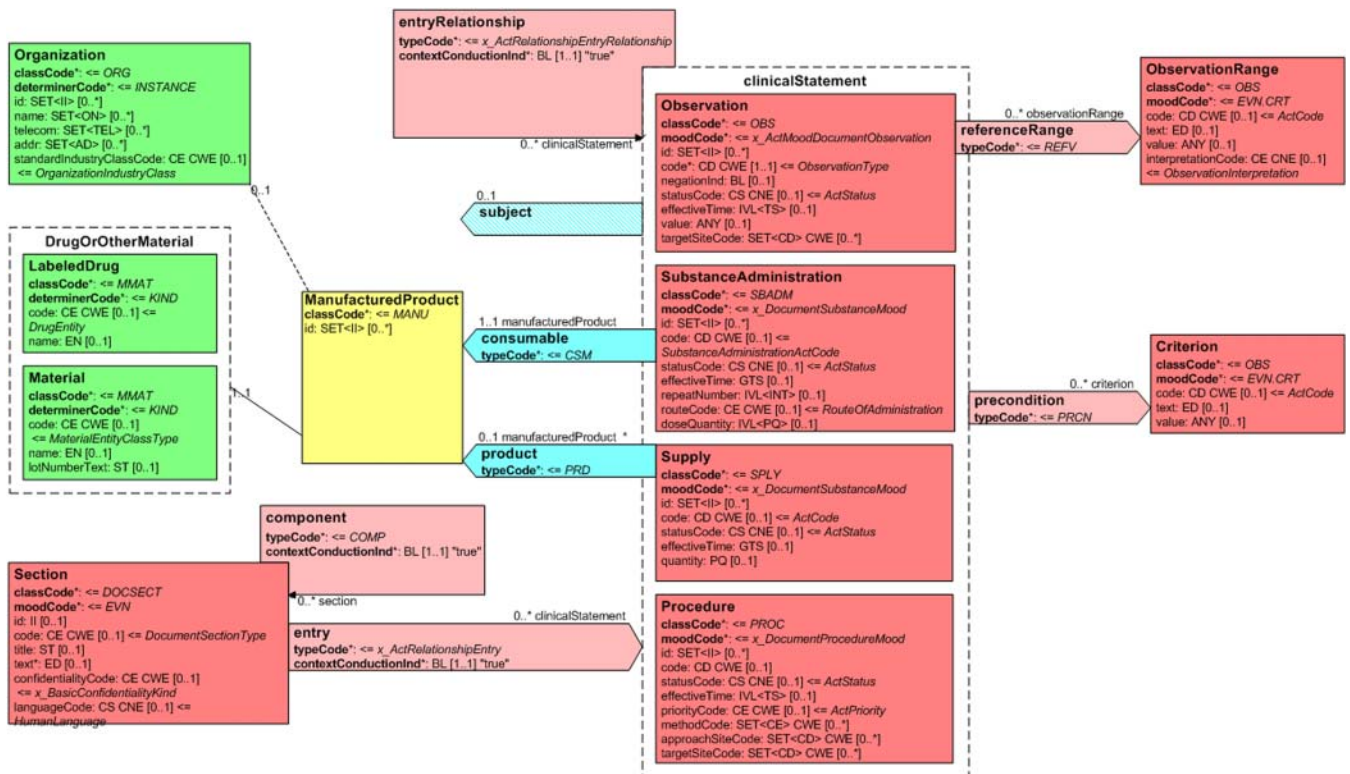


Figure 3. Clinical Document Architecture (CDA) Release 2 object model showing the connection from a document section to a portion of the CDA clinical statement model.

CDA R2 Header

The purpose of the CDA header is to set the context for the document as a whole, enable clinical document exchange across and within institutions, facilitate clinical document management, and facilitate compilation of an individual patient's clinical documents into a lifetime electronic patient record.

The entry point into the CDA model is the ClinicalDocument class, corresponding to the root <ClinicalDocument> element of the CDA Schema. The ClinicalDocument class contains various attributes such as ClinicalDocument.id, which uniquely identifies the document; ClinicalDocument.code, which specifies the particular kind of document (e.g., history and physical, discharge summary, progress note) via a code drawn from a (CWE) value set comprised of LOINC document codes³¹; ClinicalDocument.effectiveTime, which represents the time of document creation; ClinicalDocument.confidentialityCode, which defines the overall confidentiality status of the document; and ClinicalDocument.languageCode, which specifies the human language of the character data of the document.

Perhaps the most significant change to the CDA header from CDA R1 has to do with the evolution of the RIM since November 2000, which has resulted in a significant reduction in ambiguity, allowing for a more reproducible distinction between the various participants. The CDA R2 header defines many participants (such as authenticator, author, encounter Participant, legalAuthenticator, and performer). Several CDA header participants are commonly played by the same person, but this is not always the case, and the CDA R2 header adds clarity to these less common scenarios. For instance,

where a physician sees a patient as a consultant, dictates a note, and later signs it, the physician is participating as author, encounterParticipant, and legalAuthenticator. On the other hand, where a medical resident sees a patient with one staff physician, dictates a note and later signs it, and the note is co-signed by a second staff physician, the resident is author and authenticator, both the resident and the first staff physician are encounterParticipants, and the second staff physician is the legalAuthenticator.

CDA R2 document succession management operates much as it did in CDA R1. The ParentDocument Act represents a document that has been replaced or appended by the current CDA document or is the source document that was transformed into the current CDA document. Where the current CDA document is a replacement, the ParentDocument is considered superseded and is typically retained for historical or auditing purposes but no longer readily available for viewing in a patient care setting. Where the CDA document is an addendum, the ParentDocument remains a current component of the patient record, and the addendum and its Parent Document are often displayed together. A transformation is a change in the format of the ParentDocument into CDA format, and must ensure that the human readable clinical content of the report is not affected. This scenario might arise, for instance, when there is a need to exchange a document that is stored in a proprietary internal format. The document is exported and transformed into CDA format for exchange.

CDA R2 Body

Development of the model for the CDA R2 body was heavily influenced by the CEN ENV 13606, *openEHR*, and DICOM models,^{11–13} particularly in helping to determine the optimal

level of abstraction and in validating the model. The body contains the clinical report and can be either an unstructured blob, represented by the NonXMLBody class in Figure 2, or can be composed of structured markup, represented by the StructuredBody class. The NonXMLBody class is provided for those applications that can do no more than simply wrap an existing non-XML document with the CDA Header, and thus represents a very low bar for adopting the standard. The StructuredBody contains one or more Section components.

The Section class contains various attributes such as Section.id, which uniquely identifies the section; Section.code, which specifies the particular kind of section (e.g., chief complaint, allergies and adverse reactions, review of systems) via a code drawn from a (CWE) LOINC value set; Section.title represents the human readable label of a section; Section.text is the “narrative block” described above, which contains the human readable content of the section that the document originator must populate with attested content and the recipient must render using defined rendering rules.

The Section class in Figure 2 also has an author participant, modeled as a “shadow” (i.e., exact copy) of the author participant in the CDA Header. This illustrates CDA’s representation of “context,” which defines how assertions made in one part of the document propagate to or can be overridden by assertions made elsewhere in the document. Contextual components (participants, confidentiality status, language) can be set in the header, where they propagate to the body, to the sections, and to nested entries, unless overridden. In Figure 2, authorship, confidentialityCode, and language Code can be asserted in the header, and propagate through the component relationships to the body and to sections (because component.contextConductionInd is fixed as “true”). If a particular section specifies a different value for authorship, confidentiality, or language, it overrides the value propagated from the header, and the new value propagates to nested components (because contextControlCode is fixed as “OP,” which stands for Overriding and Propagating). This approach to context attempts to make explicit what humans implicitly assume when reading a document (i.e., information stated in the header is assumed to be true in the body unless stated otherwise), and obviates the need to restate all the contextual components for each section and each entry. It also leads to a fairly simple mechanism of computing the current context of any node in the CDA instance, where one only needs to walk up the XML tree and find the most proximal assertion of any contextual component, such as with this XPath expression for the author of the current node: “(ancestor-or-self::*/author)[position()=last()].”

Branching to the right of the Section class is the entry relationship, which leads to the clinical statement portion of the model, a portion of which is shown in Figure 3. CDA entries represent structured content provided for further computer processing (e.g., decision support applications). They typically encode some portion of the narrative in the Section. text field of the containing section. The dotted clinical Statement box represents a choice structure, which contains specializations of the Act class (such as Observation, SubstanceAdministration, Supply, Procedure) that provide for the formal representation.

Figures 4 through 8 illustrate various aspects of the CDA R2 clinical statement model, including how clinical statements

```
<section>
  <code code="8716-3" codeSystem="2.16.840.1.113883.6.1"
    codeSystemName="LOINC"/>
  <title>Vital Signs</title>
  <text>Temperature is 36.9 C</text>
  <entry>
    <observation classCode="OBS" moodCode="EVN">
      <code code="386725007" codeSystem="2.16.840.1.113883.6.96"
        codeSystemName="SNOMED CT" displayName="Body temperature"/>
      <statusCode code="completed"/>
      <effectiveTime value="200004071430"/>
      <value xsi:type="PQ" value="36.9" unit="Cel"/>
    </observation>
  </entry>
</section>
```

Figure 4. An example of a simple observation.

are nested within a document section. All figures are valid fragments of a CDA R2 instance, and in some cases include optional components (such as codeSystemName and displayName) to make the examples clearer.

The Observation class is used for representing coded and other observations. Figure 4 shows a simple observation, a body temperature measurement, contained within a vital signs section. It illustrates that a typical observation has an Observation.code and an Observation.value, analogous to the representation of observations in HL7 Version Two messages. The moodCode value is “EVN” (event), meaning that this observation has occurred. A status (statusCode) and time stamp (effectiveTime) are also typically expressed. Whereas HL7 V2 had a separate field to declare the data type of the observation value, HL7 V3 declares the data type of Observation.value using the xsi:type attribute. Note that the nested observation does not rely on the value of Section.code for proper interpretation and that a body temperature observation could appear elsewhere in the document, in another section.

Figure 5 is a more complex observation, a patient-reported symptom, contained within a history of present illness section. Within Section.text is a <content> element, serving as a target of an <originalText> reference from the nested observation entry. This mechanism might be used, for instance, by a natural language processing engine to associate a CDA entry with the narrative from which the entry was gleaned. The figure also illustrates the use of the CD data type to insert a qualifier, enabling post-coordination of SNOMED CT concepts to represent “osteoarthritis of the right knee.”

Acts in the clinicalStatement choice box seen in Figure 3 can relate to other Acts by traversing the entryRelationship class.

```
<section>
  <code code="10164-2" codeSystem="2.16.840.1.113883.6.1"
    codeSystemName="LOINC"/>
  <title>History of Present Illness</title>
  <text>Henry Levin, the 7<sup>th</sup> is a 67 year old male
    complaining of disabling <content ID="#SX1">osteoarthritis
    of the right knee</content>.
  </text>
  <entry>
    <observation classCode="OBS" moodCode="EVN">
      <code code="396275006" codeSystem="2.16.840.1.113883.6.96"
        codeSystemName="SNOMED CT" displayName="Osteoarthritis">
        <originalText><reference value="#SX1"/></originalText>
        <qualifier>
          <name code="363698007" codeSystem="2.16.840.1.113883.6.96"
            displayName="finding site"/>
          <value code="6757004" codeSystem="2.16.840.1.113883.6.96"
            displayName="right knee"/>
        </qualifier>
      </code>
    </observation>
  </entry>
</section>
```

Figure 5. An example of a more complex observation.


```

<section>
  <code code="10157-2" codeSystem="2.16.840.1.113883.6.1"
    codeSystemName="LOINC"/>
  <title>Family history</title>
  <text>
    <list>
      <item>Father had fatal MI in 1970.</item>
      <item>No cancer or diabetes.</item>
    </list>
  </text>
  <entry>
    <observation classCode="OBS" moodCode="EVN">
      <code code="22298006" codeSystem="2.16.840.1.113883.6.96"
        codeSystemName="SNOMED CT" displayName="Myocardial infarction"/>
      <effectiveTime value="1970"/>
      <subject>
        <relatedSubject classCode="PRS">
          <code code="FTH" codeSystem="2.16.840.1.113883.5.111"
            codeSystemName="PersonalRelationshipRoleType"
            displayName="Father"/>
        </relatedSubject>
      </subject>
      <entryRelationship typeCode="CAUS" contextConductionInd="true">
        <observation classCode="OBS" moodCode="EVN">
          <code code="399347008" codeSystem="2.16.840.1.113883.6.96"
            codeSystemName="SNOMED CT" displayName="death"/>
          <effectiveTime value="1970"/>
        </observation>
      </entryRelationship>
    </observation>
  </entry>
  <entry>
    <observation classCode="OBS" moodCode="EVN" negationInd="true">
      <code code="275937001" codeSystem="2.16.840.1.113883.6.96"
        codeSystemName="SNOMED CT"
        displayName="Family history of cancer"/>
    </observation>
  </entry>
  <entry>
    <observation classCode="OBS" moodCode="EVN">
      <code code="160274005" codeSystem="2.16.840.1.113883.6.96"
        codeSystemName="SNOMED CT"
        displayName="No family history of diabetes"/>
    </observation>
  </entry>
</section>

```

Figure 6. An example of family history observations.

The semantic relationship between the two Acts is specified in `entryRelationship.typeCode`, which has a (CNE) value set of HL7-defined codes. The CDA specification permits any Act in the `clinicalStatement` choice box to relate to any other Act using any of the enumerated relationship types, even though in many cases, this would result in nonsensical relationships. Table 1 shows a subset of the allowed relationship types, along with source and target Acts that might reasonably use them.

Figure 6 illustrates the use of the `entryRelationship` class and represents a set of family history observations, contained

```

<section>
  <code code="101155-0" codeSystem="2.16.840.1.113883.6.1"
    codeSystemName="LOINC"/>
  <title>Allergies and Adverse Reactions</title>
  <text>
    <list>
      <item>Penicillin - Hives</item>
      <item>Aspirin - Wheezing</item>
      <item>Codeine - Itching and nausea</item>
    </list>
  </text>
  <entry>
    <observation classCode="OBS" moodCode="EVN">
      <code code="247472004" codeSystem="2.16.840.1.113883.6.96"
        displayName="Hives"/>
      <entryRelationship typeCode="MFST">
        <observation classCode="OBS" moodCode="EVN">
          <code code="91936005" codeSystem="2.16.840.1.113883.6.96"
            codeSystemName="SNOMED CT" displayName="PCN Allergy"/>
        </observation>
      </entryRelationship>
    </observation>
  </entry>
</section>

```

Figure 7. An example of allergies and adverse reactions.

within a family history section. The first observation is that of a myocardial infarction (MI). The subject participant of “FTH” indicates that it was the father who suffered the MI. Traversing the `entryRelationship` class, there is a nested observation of death. The subject participant propagates to this nested relationship, and along with the `entryRelationship.typeCode` of “CAUS” indicates that MI was the cause of the father’s death. The figure also shows two different representations of negation, one via the use of the `negationInd` (for “no family history of cancer”) and the other via the use of a pre-coordinated code (for “no family history of diabetes”). One of the objectives of the HL7 TermInfo project described below is to give guidance on which format to use and/or

```

<section>
  <text>Take captopril 25mg PO every 12 hours.</text>
  <entry>
    <substanceAdministration classCode="SBADM" moodCode="RQO">
      <effectiveTime xsi:type="PIVL TS">
        <period value="12" unit="h"/>
      </effectiveTime>
      <routeCode code="PO" codeSystem="2.16.840.1.113883.5.112"
        codeSystemName="RouteOfAdministration"/>
      <doseQuantity value="1"/>
      <consumable>
        <manufacturedProduct>
          <manufacturedLabeledDrug>
            <code code="318821008" codeSystem="2.16.840.1.113883.6.96"
              codeSystemName="SNOMED CT"
              displayName="Captopril 25mg tablet"/>
          </manufacturedLabeledDrug>
        </manufacturedProduct>
      </consumable>
    </substanceAdministration>
  </entry>
</section>

```

Figure 8. An example of a substance administration.

Table 1 ■ CDA entryRelationship Types

entryRelationship. typeCode	Reasonable Source and Target Acts	Comments
CAUS (is etiology for)	[Act Observation Procedure SubstanceAdministration] CAUS [Observation]	Used to show that the source caused the target observation (for instance, source “diabetes mellitus” is the cause of target “kidney disease”).
COMP (has component)	[Act Observation Procedure SubstanceAdministration Supply] COMP [Act Observation Procedure SubstanceAdministration Supply]	Used to show that the target is a component of the source (for instance, “hemoglobin measurement” is a component of a “complete blood count”).
GEVL (evaluates (goal))	[Observation] GEVL [Observation]	Used to link an observation (intent or actual) to a goal to indicate that the observation evaluates the goal (for instance, a source observation of “walking distance” evaluates a target goal of “adequate walking distance”).
MFST (is manifestation of)	[Observation] MFST [Observation]	Used to say that the source is a manifestation of the target (for instance, source “hives” is a manifestation of target “penicillin allergy”).
RSO (has reason)	[Act Encounter Observation Procedure SubstanceAdministration Supply] RSO [Act Encounter Observation Procedure SubstanceAdministration Supply]	Used to show the reason or rationale for a service (for instance, source “treadmill test” has reason “chest pain”).
SAS (starts after start)	[Act Encounter Observation Procedure SubstanceAdministration Supply] SAS [Act Encounter Observation Procedure SubstanceAdministration Supply]	The source Act starts after the start of the target Act (for instance, source “diaphoresis” starts after the start of target “chest pain”).
SPRT (has support)	[Observation] SPRT [Observation ObservationMedia RegionOfInterest]	Used to show that the target provides supporting evidence of the source (for instance, source “possible lung tumor” has support target “mass seen on chest -x-ray”).

how to compute a transformation from one representation to another.

Figure 7 illustrates a representation of allergies and adverse reactions. The narrative block contains three items, of which one is also represented as a nested observation. That the patient has a history of hives is recorded as a distinct observation, which is then linked to another observation of penicillin allergy via an entryRelationship with typeCode of “MFST” (is manifestation of).

The SubstanceAdministration class is used for representing medication-related events such as medication history and medication administration orders. Its consumable participant is played by a LabeledDrug or Material entity in the role of a ManufacturedProduct.

Figure 8 illustrates the use of the SubstanceAdministration class. The moodCode “RQO” indicates that this is a request. The effectiveTime attribute uses the PIVL_TS (Periodic Interval of Time) data type to indicate that the act is requested to occur every 12 hours. The doseQuantity of “1” indicates that one consumable is to be administered with each act. The consumable is a “Captopril 25mg tablet.”

The Supply act is used for representing the provision of a material by one entity to another (e.g., to indicate a pharmacy request to dispense a certain quantity of pills or a certain number of refills), and coupled with a SubstanceAdministration can be used to represent a medication prescription.

Future Directions

CDA R2 represents a solid step forward in our quest for semantic interoperability, but it would be naïve to think that we have reached the end of the journey. While the framework provided by the RIM and by CDA is a critical component of

semantic interoperability, it is not sufficient, particularly given the lack of a global terminology solution, and the fact that each terminology overlaps with the RIM in different ways.³² As a result, it is possible to express a clinical statement in many ways, often with no ability for a computer to determine equivalency. Thus, while CDA R2 is highly expressive, the primary direction for the future will be to manage this potential for variability, by constraining and/or defining transformations between the allowable representations. Three activities within HL7 focusing on this next step include HL7 Templates, the HL7 Clinical Statement Model, and the HL7 Terminology project, all of which will complement CDA R2.

HL7 Templates are a constraint on a balloted model, such as the CDA R2 object model. The project goal is to provide a mechanism whereby a group such as a professional society can define best practices, which can be expressed in a standard format and implemented atop the constrained model, guiding the collection of key data elements and providing additional validation.^{33,34} Whereas CDA R2 says that a document contains sections and that sections contain observations, a template might further constrain that a particular document has particular types of sections (e.g., if ClinicalDocument.code represents a discharge summary, then there must be a nested section.code representing allergies and adverse reactions), or that a particular section contains particular types of observations (e.g., if section.code represents vital signs, then there must be a nested observation.code representing blood pressure).

The HL7 Clinical Statement Model is a collaborative project between several committees, whose focus is on harmonizing clinical statement requirements into a single model that can be used in many V3 specifications, such as CDA. The model for CDA entries was used as the starting point for and is

now derived from the shared HL7 Clinical Statement model. Sharing this model across various specifications ensures a common representation for medications, family history, signs and symptoms, etc., and fosters data reusability across HL7 V3 specifications.

The HL7 TermInfo project is jointly sponsored by the HL7 Vocabulary Committee and the College of American Pathologists and is focused on the development of implementation guidelines for the use of SNOMED CT in the HL7 Clinical Statement Model. It is anticipated that these guidelines will ultimately be balloted as an HL7 standard and jointly distributed by the College of American Pathologists as part of their SNOMED CT implementation guidance. Topics will address SNOMED CT value sets, use of the HL7 CD data type for post-coordination of SNOMED CT concepts, and aggregation of pre- vs. post-coordinated representations.

Discussion

CDA R2, being the second release of a normative ANSI-approved HL7 specification, represents a stable platform for the exchange of clinical documents. The underlying design minimizes the technical barriers to adoption while providing a migration pathway toward progressively richer computer-processable content.

Many will choose to use CDA R2 in its easy-to-implement form of just a header wrapping a non-XML body or of a header with sections containing only narrative and no entries. This will serve to bring a lot of clinical documents into a standard format. While it may be a small step, it is then possible to incrementally add structured entries. CDA R2 offers implementers the ability to use the standard now, while over time adding sophistication. An ongoing challenge facing electronic health records is user interface design so as to capture the complex information typically recorded in narrative. CDA R2 offers a rich semantic model, and it is anticipated that a better understanding of the output of a clinician interaction with an electronic health record will lead to new ideas in user interface design. If one knows the model to be populated, it stands to reason that this will lead to new ideas in data capture. Finally, the RIM and CDA R2 provide a richness of expression that may exceed the capabilities of some of today's electronic health record applications. The use of CDA R2 is anticipated to indirectly lead to enhancements in the models underlying these applications, potentially opening up new capabilities in areas such as decision support and patient safety alerts.

References ■

1. Dolin RH, Alschuler L, Beebe C, Biron PV, Boyer SL, Essin D, Kimber E, et al. The HL7 Clinical Document Architecture. *J Am Med Inform Assoc.* 2001;8:552-69.
2. Alschuler L, Dolin RH, Boyer S, Beebe C, editors. HL7 Clinical Document Architecture Framework, Release 1.0. ANSI-approved HL7 Standard, Nov 2000. Ann Arbor, MI: Health Level Seven, Inc., 2000.
3. Dolin RH, Alschuler L, Boyer S, Beebe C, Behlen FM, Biron PV, Shabo A, editors. HL7 Clinical Document Architecture, Release 2.0. ANSI-approved HL7 Standard, May 2005. Ann Arbor, MI: Health Level Seven, Inc., 2005.
4. 2nd International Conference on the Clinical Document Architecture. Oct 2004, Acapulco, Mexico. Available from: <http://www.hl7.de/iamcda2004/cdainfo.html/>. Accessed 2005 Jun 11.
5. Poulymenopoulou M, Vassilacopoulos G. An electronic patient record implementation using clinical document architecture. *Stud Health Technol Inform.* 2004;103:50-7.
6. Guo J, Takada A, Tanaka K, Sato J, Suzuki M, Suzuki T, et al. The development of MML (Medical Markup Language) version 3.0 as a medical document exchange format for HL7 messages. *J Med Syst.* 2004;28:523-33.
7. Heitmann KU, Schweiger R, Dudeck J. Discharge and referral data exchange using global standards—the SCIPHOX project in Germany. *Stud Health Technol Inform.* 2002;90:679-84.
8. Rishel W. A new approach to claims attachments. *Healthcare Inform.* 2003. Available from: http://www.healthcare-informatics.com/issues/2003/09_03/rishel.htm/. Accessed 2005 Jun 11.
9. HL7 Claims Attachments Implementation Guides. © Health Level Seven, Aug 2003.
10. Huang Y, Lowe HJ, Klein D, Cucina RJ. Improved identification of noun phrases in clinical radiology reports using a high-performance statistical natural language parser augmented with the UMLS specialist lexicon. *J Am Med Inform Assoc.* 2005;12:275-85.
11. European Committee for Standardization, Technical Committee for Health Informatics (CEN TC251). Four-part EHCR Message Standard. Publication CEN ENV 13606. Brussels, Belgium: CEN TC251.
12. The openEHR Foundation. Available from: <http://www.openehr.org/>.
13. National Electrical Manufacturers Association. Digital Imaging and Communications in Medicine (DICOM), supplement 23: structured reporting. Rosslyn, VA: NEMA, 1999.
14. Friedman C, Hripcsak G, Shagina L, Liu H. Representing information in patient reports using natural language processing and the extensible markup language. *J Am Med Inform Assoc.* 1999; 6:76-87.
15. World Wide Web Consortium (W3C). Extensible Markup Language (XML) 1.0. Recommendation, Feb 10, 1998. Available from: <http://www.w3.org/TR/1998/REC-xml-19980210.html/>. Accessed 2005 Jun 11.
16. Health Level Seven, Inc. HL7 Reference Information Model. Ann Arbor, MI: Health Level Seven, Inc., 1994. Available from: http://www.hl7.org/library/data-model/RIM/modelpage_non.htm. Accessed 2005 Jun 11.
17. Bakken S, Campbell KE, Cimino JJ, Huff SM, Hammond WE. Toward vocabulary domain specifications for Health Level 7-coded data elements. *J Am Med Inform Assoc.* 2000;7:333-42.
18. SNOMED Clinical Terms. College of American Pathologists. Available from: <http://www.snomed.org/>. Accessed 2005 Jun 11.
19. Logical Observation Identifiers Names and Codes (LOINC). Available from: <http://www.loinc.org/>. Accessed 2005 Jun 11.
20. World Wide Web Consortium (W3C). XHTML 1.0: The Extensible HyperText Markup Language—A Reformulation of HTML 4 in XML 1.0. Recommendation 2000 Jan 26. Available from: <http://www.w3.org/TR/xhtml1/>. Accessed 2005 Jun 11.
21. MIME Encapsulation of Aggregate Documents, such as HTML (MHTML). Copyright © The Internet Society (1999). Available from: <http://www.ietf.org/rfc/rfc2557.txt>. Accessed 2005 Jun 11.
22. Beeler GW Jr. HL7 version 3: an object-oriented methodology for collaborative standards development. *Int J Med Inform.* 1998;48: 151-61.
23. Beeler GW Jr. Taking HL7 to the next level. *MD Comput.* 1999; 16:21-4.
24. Dolin RH. Advances in data exchange for the clinical laboratory. *Clin Lab Med.* 1999;19:385-419.
25. Health Level Seven. HL7 Version 3 Development Framework (HDF). Ann Arbor, MI: Health Level Seven, Inc., 2005.
26. World Wide Web Consortium (W3C). XML Schema Part 1: Structures. W3C working draft, May 6, 1999. Available from: <http://www.w3.org/TR/xmlschema-1/>.

- www.w3.org/1999/05/06-xmlschema-1/. Accessed 2005 Jun 11.
27. World Wide Web Consortium (W3C). XML Schema Part 2: Datatypes. W3C working draft, May 6, 1999. Available from: <http://www.w3.org/1999/05/06-xmlschema-2/>. Accessed 2005 Jun 11.
 28. Dolin RH, Spackman KA, Markwell D. Selective retrieval of pre- and post-coordinated SNOMED concepts. JAMIA Fall Symp Suppl. 2002;210-4.
 29. Brown PJB, Sonksen P. Evaluation of the quality of information retrieval of clinical findings from a computerized patient database using a semantic terminological model. J Am Med Inform Assoc. 2000;7:392-403.
 30. Unified Modeling Language. Available from: <http://www.uml.org/>. Accessed 2005 Jun 11.
 31. Frazier P, Rossi-Mori A, Dolin RH, Alschuler L, Huff SM. The creation of ontology of clinical document names. Medinfo. 2001;10:94-8.
 32. Rector AL. The interface between information, terminology, and inference models. Medinfo. 2001;10:246-50.
 33. Mattison JE, Dolin RH, Laberge D. Managing the tensions between national standardization vs. regional localization of clinical content and templates. Medinfo. 2004;11:1081-5.
 34. Shabo A. Structuring the medical narrative in patient records—a further step towards a multi-accessible EHR. Yearbook Med Inform. 2004;503-6.